B1

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/00, C12P 21/00, 21/02 — A1

(11) International Publication Number: WO 90/10697

(43) International Publication Date: 20 September 1990 (20.09.90)

(21) International Application Number: PCT/US90/01353

(22) International Filing Date: 15 March 1990 (15.03.90)

(30) Priority data:
323,964 15 March 1989 (15.03.89) US

(60) Parent Application or Grant
(63) Related by Continuation
US 323,964 (CIP)
Filed on 15 March 1989 (15.03.89)

(71) Applicant *(for all designated States except US)*: THE SALK INSTITUTE BIOTECHNOLOGY/INDUSTRIAL ASSOCIATES, INC. [US/US]; 10280 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)* : SIEGEL, Robert, Steven [US/US]; 2633 Wyandotte, San Diego, CA 92117 (US). BUCKHOLZ, Richard, Gordon [US/US]; 1667 Clearwater Place, Encinitas, CA 92024 (US). THILL, Gregory, Patrick [US/US]; 3284 West Fox Run Way, San Diego, CA 92111 (US). WONDRACK, Lillian, Margaret [US/US]; 1527 Grand Avenue, San Diego, CA 92109 (US).

(74) Agents: WATT, Phillip, H. et al.; Fitch, Even, Tabin & Flannery, Room 900, 135 South LaSalle Street, Chicago, IL 60603 (US).

(81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: PRODUCTION OF EPIDERMAL GROWTH FACTOR IN METHYLOTROPHIC YEAST CELLS

(57) Abstract

Epidermal growth factor, a naturally occurring, relatively short, single chain polypeptide, is prepared by growing methylotrophic yeast transformants containing in their genome at least one copy of a DNA sequence operably encoding EGF, in operational association with a DNA sequence encoding the S. cerevisiae alpha mating factor pre-pro sequence (including the proteolytic processing site: lys-arg), both under the regulation of a promoter region of a gene of a methylotrophic yeast, under conditions allowing expression of said DNA sequences, and secretion of EGF into the culture medium. Also disclosed are novel DNA fragments and novel recombinant yeast strains which are useful in the practice of the present invention.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MR	Mauritania
BE	Belgium	GA	Gabon	MW	Malawi
BF	Burkina Faso	GB	United Kingdom	NL	Netherlands
BG	Bulgaria	HU	Hungary	NO	Norway
BJ	Benin	IT	Italy	RO	Romania
BR	Brazil	JP	Japan	SD	Sudan
CA	Canada	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic			SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

PRODUCTION OF EPIDERMAL GROWTH FACTOR IN METHYLOTROPHIC YEAST CELLS

Field of the Invention

This invention relates to a process of recombinant DNA technology for producing epidermal growth factor (EGF) peptides in methylotrophic yeast such as Pichia pastoris. Methylotrophic yeast transformants containing in their genome at least one copy of a DNA sequence operably encoding an EGF peptide under the regulation of a promoter region of a gene of a methylotrophic yeast and the S. cerevisiae alpha-mating factor (AMF) pre-pro sequence are cultured under conditions allowing the expression of EGF peptides into the culture medium. The invention further relates to the methylotrophic yeast transformants, DNA fragments and expression vectors used for their production and cultures containing same.

Background of the Invention

Epidermal growth factor (EGF) is a naturally-occurring, relatively short, single-chain polypeptide, which was first isolated from the mouse submaxillary gland. A structurally very similar polypeptide was later detected and isolated from human urine at low (about 30 ng/ml) concentrations. Both mouse and human epidermal growth factors (the latter also called urogastrone in some earlier publications) contain 53 amino acids. Thirty-seven of these are identical in the amino acid

sequences of mouse epidermal growth factor (mEGF) and human epidermal growth factor (hEGF), as are the relative positions of the three disulfide bonds present in the structure. [Gregory, Nature, 257, 325 (1975); Gregory et al., Hoppe-Seyler's Z. Physiol. Chem., 356, 1765 (1975)]. The amino acid sequence of the form of hEGF containing 53 amino acids (β-hEGF), as reported in the literature, is as follows:

Asn Ser Asp Ser Glu Cys Pro Leu Ser
His Asp Gly Tyr Cys Leu His Asp Gly
Val Cys Met Tyr Ile Glu Ala Leu Asp
Lys Tyr Ala Cys Asn Cys Val Val Gly
Tyr Ile Gly Glu Arg Cys Gln Tyr Arg
Asp Leu Lys Trp Trp Glu Leu Arg

The polypeptide also exists as a 52 amino acid form (gamma-hEGF) that lacks the C-terminal arginine residue found in β-hEGF.

The amino acid and nucleotide sequences of hEGF are, for example, disclosed in Hollenberg, "Epidermal Growth Factor-Urogastrone, A Polypeptide Acquiring Hormonal States"; eds., Academic Press, Inc., New York (1979), pp. 69-110; or Urdea et al., Proc. Natl. Acad. Sci. USA, 80, 7461 (1983).

A 48 amino acid-containing form of hEGF (lacking the five C-terminal amino acids) is described in Japanese Patent Application 86146964, published 8 February 1988 under No. 63003791.

The molecule in natural form contains disulfide linkages between residues 6-20, 14-31 and 33-42, and arises from an about 1200 amino acid precursor molecule consisting of eight EGF-like regions [see e.g. Bell et al., Nucleic Acid Research, 14, 21, 8427 (1986)]. A form of rat EGF containing 48 amino acids has recently been disclosed in the Japanese Patent Application 8736498, published 22 August 1988, under No. 63202387. Both mEGF and hEGF, as well as their known analogs, exhibit similar pharmacological activities, although the extent or

spectrum of activity may be different for different materials. In general EGF inhibits the secretion of gastric acid and promotes cell growth; therefore, it is targeted for therapeutic potential as, for example, an anti-ulcer agent and in external wound healing.

Since isolation from natural sources is technically difficult, expensive, and time consuming, recent efforts have centered on the development of efficient recombinant methods for the production of EGF.

Of the hosts widely used for the production of heterologous proteins, probably E. coli and Saccharomyces cerevisiae (Baker's yeast) are the best understood. However, E. coli tends to produce EGF in its reduced form which is not stable in the presence of endogenous bacterial proteases. Attempts to overcome this problem, e.g., by employing a suitable leader sequence in order to produce an insoluble fusion protein which can be readily recovered from the cell paste, resulted in other inconveniences, especially during purification of the product.

Yeasts can offer clear advantages over bacteria in the production of heterologous proteins, which include their ability to secrete heterologous proteins into the culture medium. Secretion of proteins from cells is generally superior to production of proteins in the cytoplasm. Secreted products are obtained in a higher degree of initial purity and their further purification is easier to contend with without cellular debris. In the case of sulfhydryl-rich proteins there is another compelling reason for the development of hosts capable of secreting them into the culture medium: their correct tertiary structure is produced and maintained via disulfide bonds. The secretory pathway of the cell and the extracellular medium are oxidizing environments which can support disulfide bond formation [Smith, et al., Science, 229, 1219 (1985)]. In contrast, the cytoplasm is a reducing environment in which disulfide bonds cannot

form. Upon cell breakage, too rapid formation of disulfide linkages can result in random disulfide bond formation. Consequently, production of sulfhydryl rich proteins, such as EGF, containing appropriately formed disulfide bonds can be best achieved by transit through the secretory pathway.

Secretion of authentic biologically active human epidermal growth factor from S. cerevisiae is disclosed in European Patent Application Nos. 84104445.6 and 84303783.9, published October 31, 1984 (No. 0 123 289) and December 19, 1984 (No. 0 128 733), respectively. The cited patent applications contain no details as to the level of secretion or the purity of hEGF obtained. In an article published in Proc. Natl. Acad. Sci. USA, 81, 4642 (1984), Brake, inventor of European Patent Application No. 84104445.6, and his co-workers give more details of their laboratory-scale experiments. hEGF is produced in S. cerevisiae by means of an expression cassette containing a DNA sequence encoding mature hEGF joined to sequences encoding the leader region ("pre-pro" segment) of the precursor of the yeast mating pheromone alpha-factor. In what appears to be the best experiment, hEGF was secreted into the shake flask culture medium in a concentration of about 4000 ng/ml. In view of the problems usually encountered with up-scaling the production of heterologous proteins in autonomous plasmid-based yeast systems, such as S. cerevisiae, there is no indication that hEGF production in S. cerevisiae could be at levels higher than those of that experimental system.

According to the prior art methods hEGF is produced and secreted from yeast in mature form, usually containing 52 amino acids.

To overcome the major problems associated with S. cerevisiae, e.g. loss of selection for plasmid maintenance and problems concerning plasmid distribution, copy number and stability in fermentors operated at high

cell density, a yeast expression system based on the methylotrophic yeast Pichia pastoris has been developed. A key feature making this system unique lies with the promoter employed to drive heterologous gene expression. This promoter, which is derived from the methanol-regulated alcohol oxidase I (AOX1) gene of P. pastoris, is highly expressed and tightly regulated (see e.g. the European Patent Application No. 85113737.2, published June 4, 1986, under No. 0 183 071). Another key feature of the P. pastoris expression system is the stable integration of expression cassettes into the P. pastoris genome, thus significantly decreasing the chance of vector loss.

Although P. pastoris has been used successfully for the production of various heterologous proteins, e.g., hepatitis B surface antigen [Cregg et al., Bio/Technology 5, 479 (1987)], lysozyme and invertase [Digan et al., Developments in Industrial Microbiology 29, 59 (1988); Tschopp et al., Bio/Technology 5, 1305 (1987)], endeavors to produce other heterologous gene products in Pichia, especially by secretion, have given mixed results. At our present level of understanding of the P. pastoris expression system, it is unpredictable whether a given gene can be expressed to an appreciable level in this yeast or whether Pichia will tolerate the presence of the recombinant gene product in its cells. Further, it is especially difficult to foresee if a particular protein will be secreted by P. pastoris, and if it is, at what efficiency. Even for S. cerevisiae, which has been considerably more extensively studied than P. pastoris, the mechanism of protein secretion is not well defined and understood.

Summary of the Invention

The present invention provides an expression system suitable for the production of EGF. In addition,

the present invention provides a powerful method for the production of secreted EGF peptides in methylotrophic yeast such as Pichia pastoris, which method can be easily scaled up from shake-flask cultures to large fermentors with no loss in productivity and without making major changes in the fermentation conditions. The presently preferred yeast species for use in the practice of the present invention is Pichia pastoris, a known industrial yeast strain that is capable of utilizing methanol as the sole carbon and energy source (methylotroph). We have surprisingly found that EGF peptides can be produced in and secreted from methylotrophic yeast such as P. pastoris very efficiently, by transforming a methylotrophic yeast with, and preferably integrating into the yeast genome, at least one copy of a first DNA sequence operably encoding an EGF peptide, wherein said first DNA sequence is operably associated with a second DNA sequence encoding the S. cerevisiae alpha-mating factor (AMF) pre-pro sequence (including the proteolytic processing site: lys-arg), and wherein both of said DNA sequences are under the regulation of a methenol responsive promoter region of a gene of a methylotrophic yeast. Methylotrophic yeast cells such as P. pastoris cells containing in their genome at least one copy of these DNA sequences efficiently produce biologically active EGF peptides as a medium secreted product.

Accordingly, this invention relates to a methylotrophic yeast cell such as a P. pastoris cell containing in its genome at least one copy of a DNA sequence operably encoding an EGF peptide, operably associated with a DNA sequence encoding the S. cerevisiae AMF pre-pro sequence (including the proteolytic processing site: lys-arg), both under the regulation of a promoter region of a gene of a methylotrophic yeast.

According to another aspect, this invention relates to a DNA fragment containing at least one copy of

an expression cassette comprising in the reading frame direction of transcription, the following DNA sequences:

(i) a promoter region of a methanol responsive gene of a methylotrophic yeast,

(ii) a DNA sequence encoding a polypeptide consisting of:

(a) the S. cerevisiae AMF pre-pro sequence, including the proteolytic processing site: lys-arg, and

(b) a DNA sequence encoding an EGF peptide; and

(iii) a transcription terminator functional in a methylotrophic yeast,

wherein said DNA sequences are operationally associated with one another for transcription of the sequences encoding said polypeptide.

The DNA fragment according to the invention can be transformed into the methylotrophic yeast cells such as P. pastoris cells as a linear fragment flanked by DNA sequences having sufficient homology with a target gene to effect integration of said DNA fragment therein. In this case integration takes place by replacement at the site of the target gene. Alternatively, the DNA fragment can be part of a circular plasmid, which may be linearized to facilitate integration, and will integrate by addition at a site of homology between the host and the plasmid sequence.

The invention further concerns an expression vector containing at least one copy of an expression cassette described hereinabove.

According to a still further embodiment, the invention relates to a process for producing EGF peptides by growing methylotrophic yeast transformants containing in their genome at least one copy of a DNA sequence operably encoding an EGF peptide, operably associated with DNA encoding the S. cerevisiae AMF pre-pro sequence, both under the regulation of a promoter region of a

conditions allowing the expression of said DNA sequence in said transformants and secreting mature EGF peptides into the culture medium. Cultures of viable methylotrophic yeast cells capable of producing EGF peptides are also within the scope of the invention.

The polypeptide product is secreted to the culture medium at surprisingly high concentrations; the level of EGF peptides secretion is about two orders of magnitude higher than the best results published in the literature. In addition to the unique properties of the invention expression system, these present, excellent results are also due to the fact that the S. cerevisiae alpha-mating factor pre-pro sequence functions unexpectedly well to direct secretion of EGF peptides in methylotrophic yeast such as P. pastoris.

Another surprising discovery is that the full length, 1-52 form of hEGF secreted by P. pastoris cells is not stable in the broth; it gets degraded to a shorter 1-48 amino acid containing, stable form. The shorter hEGF form has essentially the same biological activity as the full length hEGF.

The present invention is directed to the above aspects and all associated methods and means for accomplishing such. For example, the invention includes the technology requisite to suitable growth of the methylotrophic yeast host cells, fermentation, and isolation and purification of the EGF gene product.

P. pastoris is described herein as a model system for the use of methylotrophic yeast hosts. Other useful methylotrophic yeasts can be taken from four genera, namely Candida, Hanensula, Pichia and Torulopsis. Equivalent species from them may be used as hosts herein primarily based upon their demonstrated characterization of being supportable for growth and exploitation on methanol as a single carbon nutriment source. See, for example, Gleeson et al., Yeast 4, 1 (1988).

Brief Description of the Drawings

Figure 1 shows the restriction map and insert sequence of the hEGF gene employed herein.

Figure 2 shows the restriction map and insert sequence of the AMF pre-pro fragment (including the proteolytic processing site: lys-arg) employed herein.

Figure 3 is a restriction map of plasmid pAO208.

Figure 4 is a restriction map of plasmid pAO817.

Figure 5 is a restriction map of plasmid pEGF819.

Detailed Description of the Invention

The term "epidermal growth factor" or "EGF peptide" or simply "EGF", as used throughout the specification and in the claims, refers to a polypeptide product which exhibits similar, in-kind, biological activities to natural human epidermal growth factor (hEGF), as measured in recognized bioassays, and has substantially the same amino acid sequence as hEGF, including the 53, 52 and 48 amino acid forms. It will be understood that polypeptides deficient in one or more amino acids in the amino acid sequence reported in the literature for naturally occurring hEGF, or polypeptides containing additional amino acids or polypeptides in which one or more amino acids in the amino acid sequence of natural hEGF are replaced by other amino acids are within the scope of the invention, provided that they exhibit the functional activity of hEGF, e.g., inhibition of the secretion of gastric acid and promotion of cell growth. The invention is intended to embrace all the allelic variations of hEGF. Moreover, as noted Supra, derivatives obtained by simple modification of the amino acid sequence of the naturally occurring product, e.g, by

way of site-directed mutagenesis or other standard procedures, are included within the scope of the present invention. EGF forms produced by proteolysis of host cells that exhibit similar biological activities to mature, naturally occurring hEGF are also encompassed by the present invention.

The amino acids, which occur in the various amino acid sequences referred to in the specification have their usual, three- and one-letter abbreviations, routinely used in the art, i.e.:

Amino Acid	Abbreviation	
L-Alanine	Ala	A
L-Arginine	Arg	R
L-Asparagine	Asn	N
L-Aspartic acid	Asp	D
L-Cysteine	Cys	C
L-Glutamine	Gln	Q
L-Glutamic Acid	Glu	E
L-Glycine	Gly	G
L-Histidine	His	H
L-Isoleucine	Ile	I
L-Leucine	Leu	L
L-Lysine	Lys	K
L-Methionine	Met	M
L-Phenylalanine	Phe	F
L-Proline	Pro	P
L-Serine	Ser	S
L-Threonine	Thr	T
L-Tryptophan	Trp	W
L-Tyrosine	Tyr	Y
L-Valine	Val	V

According to the invention, EGF peptides are produced by methylotrophic yeast cells containing in their genome at least one copy of a DNA sequence operably encoding EGF peptides operably associated with DNA encoding the S. cerevisiae α-mating factor (AMF) pre-pro sequence (including the proteolytic processing site: lys-arg), both under the regulation of a promoter region of a methanol responsive gene of a methylotrophic yeast.

The term "a DNA sequence operably encoding EGF peptides" as used herein includes DNA sequences encoding the 53, 52 and 48 amino acid forms of hEGF or any other

"EGF peptide" as defined herein above. DNA sequences encoding EGF, e.g. hEGF, are known in the art. They may be obtained by chemical synthesis or by transcription of a messenger RNA (mRNA) corresponding to EGF to a complementary DNA (cDNA) and converting the latter into a double stranded cDNA. The mRNA can be isolated for example, from adult mouse kidney [Rall et al., Nature, 313, 228 (1985)] or from adult human kidney (Bell et al., Nucleic Acid Research, 14, 21, 8427 (1986)]. Chemical synthesis of a gene for human EGF is, for example, disclosed by Urdea et al., Supra. The requisite DNA sequence can also be removed, for example, by restriction enzyme digest of known vectors harboring the EGF gene. Examples of such vectors and the means for their preparation can be taken from the following publications: Brake et al., Supra - e.g. the pBR322-based vector pYα EGF-21; Urdea et al., Supra - plasmid pYEGF-2, etc. The structure of a preferred hEGF gene used in accordance with the present invention is further elucidated in the examples.

The presently preferred promoter region employed to drive the EGF gene expression is derived from a methanol-regulated alcohol oxidase gene of P. pastoris. P. pastoris is known to contain two functional alcohol oxidase genes: alcohol oxidase I (AOX1) and alcohol oxidase II (AOX2) genes. The coding portions of the two AOX genes are closely homologous at both the DNA and the predicted amino acid sequence levels and share common restriction sites. The proteins expressed from the two genes have similar enzymatic properties but the promoter of the AOX1 gene is more efficient and highly expressed, therefore, its use is preferred for EGF expression. The AOX1 gene, including its promoter, has been isolated and thoroughly characterized [Ellis et al., Mol. Cell. Biol. 5, 1111 (1985)].

The expression cassette used for transforming methylotrophic yeast cells contains, in addition to a

methenol responsive promoter of a methylotrophic yeast gene and the EGF encoding DNA sequence (EGF gene), a DNA sequence encoding the in-reading frame S. cerevisiae AMF pre-pro sequence, including a DNA sequence encoding the processing site: lys-arg (also referred to as the lys-arg encoding sequence), and a transcription terminator functional in a methylotrophic yeast.

The S. cerevisiae alpha-mating factor is a 13-residue peptide, secreted by cells of the "alpha" mating type, that acts on cells of the opposite "a" mating type to promote efficient conjugation between the two cell types and thereby formation of "a-alpha" diploid cells [Thorner et al., The Molecular Biology the Yeast Saccharomyces, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 143 (1981)]. The AMF pre-pro sequence is a leader sequence contained in the AMF precursor molecule, and includes the lys-arg encoding sequence which is necessary for proteolytic processing and secretion (see e.g. Brake et al., Supra). The AMF pre-pro sequence, including the lys-arg encoding sequence is a 255 bp fragment which is shown in Figure 2.

The transcription terminator functional in a methylotrophic yeast used in accordance with the present invention has a subsegment which encodes a polyadenylation signal and polyadenylation site in the transcript and/or a subsegment which provides a transcription termination signal for transcription from the promoter used in the expression cassette according to the invention (the term "expression cassette" as used herein and throughout the specification and claims refers to a DNA sequence which includes sequences functional for expression and the secretion processes). The entire transcription terminator is taken from a protein-encoding gene, which may be the same or different from the gene which is the source of the promoter used according to the invention.

For the practice of the present invention it is preferred that multiple copies of the above-described expression cassettes be contained on one DNA fragment, preferably in a head-to-tail orientation. It is particularly preferred that four or more copies of the above-described expression cassette be contained on one DNA fragment.

The DNA fragments according to the invention optionally further comprise a selectable marker gene. For this purpose, any selectable marker gene functional in methylotrophic yeast such as P. pastoris may be employed, i.e., any gene which confers a phenotype upon methylotrophic yeast cells such as P. pastoris cells thereby allowing them to be identified and selectively grown from among a vast majority of untransformed cells. Suitable selectable marker genes include, for example, selectable marker systems composed of an auxotrophic mutant P. pastoris host strain and a wild type biosynthetic gene which complements the host's defect. For transformation of his4$^-$ P. pastoris strains, for example, the S. cerevisiae or P. pastoris HIS4 gene, or for transformation of arg4$^-$ mutants the S. cerevisiae ARG4 gene or the P. pastoris ARG4 gene, may be employed.

If the yeast host is transformed with a linear DNA fragment containing the EGF gene under the regulation of a promoter region of a P. pastoris gene and AMF sequences necessary for processing and secretion, the expression cassette is integrated into the host genome by any of the gene replacement techniques known in the art, such as by one-step gene replacement [see e.g., Rothstein, Methods Enzymol. 101, 202 (1983); Cregg et al., Bio/Technology 5, 479 (1987)] or by two-step gene replacement methods [see e.g., Scherer and Davis, Proc. Natl. Acad. Sci. USA, 76, 4951 (1979)]. The linear DNA fragment is directed to the desired locus, i.e., to the target gene to be disrupted, by means of flanking DNA sequences having sufficient homology with the target gene

to effect integration of the DNA fragment therein. One-step gene disruptions are usually successful if the DNA to be introduced has as little as 0.2 kb homology with the fragment locus of the target gene; it is however, preferable to maximize the degree of homology for efficiency.

If the DNA fragment according to the invention is contained within or is an expression vector, e.g., a circular plasmid, one or more copies of the plasmid can be integrated at the same or different loci, by addition to the genome instead of by gene disruption. Linearization of the plasmid by means of a suitable restriction endonuclease facilitates integration.

The term "expression vector" includes vectors capable of expressing DNA sequences contained therein, where such sequences are in operational association with other sequences capable of effecting their expression, i.e., promoter sequences. In general, expression vectors usually used in recombinant DNA technology are often in the form of "plasmids", i.e., circular, double-stranded DNA loops which in their vector form, are not bound to the chromosome. In the present specification the terms "vector" and "plasmid" are used interchangeably. However, the invention is intended to include other forms of expression vectors as well, which function equivalently.

In the DNA fragment according to the invention the segments of the expression cassette are "operationally associated" with one another. The DNA sequence encoding EGF peptides is positioned and oriented functionally with respect to the promoter, the DNA sequence encoding the S. cerevisiae AMF pre-pro sequence (including the DNA sequence encoding the AMF processing-site: lys-arg), and the transcription terminator. Thus, the polypeptide encoding segment is transcribed, under regulation of the promoter region, into a transcript capable of providing, upon translation, the desired

polypeptide. Because of the presence of the AMF pre-pro sequence, the expressed EGF product is found as a secreted entity in the culture medium. Appropriate reading frame positioning and orientation of the various segments of the expression cassette are within the knowledge of persons of ordinary skill in the art; further details are given in the Examples.

The DNA fragment provided by the present invention may include sequences allowing for its replication and selection in bacteria, especially E. coli. In this way, large quantities of the DNA fragment can be produced by replication in bacteria.

Methods of transforming methylotrophic yeast such as Pichia pastoris as well as methods applicable for culturing methylotrophic yeast such as P. pastoris cells containing in their genome a gene for a heterologous protein are known generally in the art.

According to the invention, the expression cassettes are transformed into the cells of a methylotrophic yeast either by the spheroplast technique, described by Cregg et al., Mol. Cell. Biol. 5, 3376 (1985) or by the whole-cell lithium chloride yeast transformation system [Ito et al., Agric. Biol. Chem. 48, 341 (1984)], with modification necessary for adaptation to P. pastoris [See EP 312, 934]. Although the whole-cell lithium chloride method is more convenient in that it does not require the generation and maintenance of spheroplasts, for the purpose of the present invention the spheroplast method is preferred, primarily since it yields a greater number of transformants.

Positive transformants are characterized by Southern blot analysis [Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (1982)] for the site of DNA integration, Northern blots [Maniatis, Op. Cit., R.S. Zitomer and B.D. Hall, J. Biol. Chem, 251, 6320 (1976)] for methanol-responsive EGF gene

expression, and product analysis for the presence of secreted EGF peptides in the growth media.

Transformed strains, which are of the desired phenotype and genotype, are grown in fermentors. For the large-scale production of recombinant DNA-based products in methylotrophic yeast such as P. pastoris, a three-stage, high cell-density, batch fermentation system is normally employed. In the first, or growth stage, expression hosts are cultured in defined minimal medium with excess glycerol as carbon source. When grown on this carbon source heterologous gene expression is completely repressed, which allows the generation of cell mass in the absence of heterologous protein expression. Next, a short period of glycerol limitation growth is allowed. Subsequent to the glycerol limited growth, methanol is added, initiating the expression of the desired heterologous protein. This third stage is the so-called production stage.

The term "culture" means a propagation of cells in a medium conducive to their growth, and all sub-cultures thereof. The term "subculture" refers to a culture of cells grown from cells of another culture (source culture), or any subculture of the source culture, regardless of the number of subculturings which have been performed between the subculture of interest and the source culture.

According to a preferred embodiment of the invention, the heterologous protein expression system used for EGF production utilizes the promoter derived from the methanol-regulated AOX1 gene of P. pastoris, which is very efficiently expressed and tightly regulated. This gene can be the source of the transcription terminator as well. The presently preferred expression cassette comprises, operationally associated with one another, a P. pastoris AOX1 promoter, DNA encoding the S. cerevisiae AMF pre-pro sequence (including the DNA sequence encoding the AMF processing

site: lys-arg), a DNA sequence encoding mature hEGF, and a transcription terminator derived from the P. pastoris AOX1 gene. Preferably, two or more of such expression cassettes are contained on one DNA fragment, in head-to-tail orientation, to yield multiple expression cassettes on a single contiguous DNA fragment.

The presently preferred host cells to be transformed with multiple expression cassettes are P. pastoris cells having at least one mutation that can be complemented with a marker gene present on a transforming DNA fragment. Preferably his4$^-$ (GS115) or arg4$^-$ (GS190) auxotrophic mutant P. pastoris strains are employed.

The fragment containing multiple expression cassettes is inserted into a plasmid containing a marker gene complementing the host's defect. pBR322-based plasmids, e.g., pAO815, are preferred. Insertion of multiple copies of the hEGF expression/secretion cassette into parent plasmid pAO815 produces plasmids pAO817 and pEGF819.

To develop Mut$^-$ expression strains of P. pastoris, the transforming DNA comprising the expression cassette(s) is (are) preferably integrated into the host genome by a one-step gene replacement technique. The expression vector is digested with an appropriate enzyme to yield a linear DNA fragment with ends homologous to the AOX1 locus by means of the flanking homologous sequences. This approach avoids the problems encountered with S. cerevisiae, wherein expression cassettes must be present on multicopy plasmids to achieve high level of expression. As a result of gene replacement, Mut$^-$ strains are obtained. Mut refers to the methanol-utilization phenotype. In Mut$^-$ strains, the AOX1 gene is replaced with the expression cassette(s), thus decreasing the strains ability to utilize methanol. A slow growth rate on methanol is maintained by expression of the AOX2 gene product. The transformants in which the expression cassette has integrated into the AOX1 locus by site-

directed recombination can be identified by first screening for the presence of the complementing gene. This is preferably accomplished by growing the cells in a media lacking the complementing gene product and identifying those cells which are able to grow by nature of expression of the complementing gene. Next, the selected cells are screened for their Mut phenotype by growing them in the presence of methanol and monitoring their growth rate.

To develop Mut$^+$ EGF-expressing strains, the fragment comprising one or more expression cassette(s) preferably is integrated into the host genome by transformation of the host with a circular plasmid comprising the expression cassette(s). The integration is by addition at a locus or loci having homology with one or more sequences present on the transformation vector.

Positive transformants are characterized by Southern analysis for the site of DNA integration, by Northern analysis for methanol-responsive EGF gene expression, and by product analysis for the presence of secreted hEGF peptides in the growth media. Methylotrophic yeast strains which have integrated one or multiple copies of the expression cassettes at a desired site can be identified by Southern blot analysis. Strains which demonstrate enhanced secretion of hEGF may be identified by Northern or product analysis; however, this characteristic is not always easy to detect in shake-flask experiments.

Methylotrophic yeast transformants which are identified to have the desired genotype and phenotype are grown in fermentors. Typically a three-step production process is used. Initially, cells are grown on a repressing carbon source, preferably excess glycerol. In this stage the cell mass is generated in absence of expression. Next, a short period of glycerol limitation growth is allowed. After exhaustion of glycerol,

methanol alone (methanol excess fed-batch mode) or limiting glycerol and methanol (mixed-feed fed-batch mode) are added in the fermentor, resulting in the expression of the hEGF gene driven by a methanol responsive promoter. The level of hEGF secreted into the media can be determined by Western blot analysis of the media in parallel with an EGF standard, using anti-EGF antisera, or by HPLC after suitable pretreatment of the medium.

The invention is further illustrated by the following non-limiting examples.

Examples

Example 1

The expression vector constructions disclosed in the present application were performed using standard procedures, as described, for example in Maniatis et al., Supra, and Davis et al., Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York (1986).

The hEGF gene was obtained from a pBR322-based plasmid on an NcoI-HindIII fragment. The hEGF encoding fragment employed is shown in Figure 1.

The AMF pre-pro encoding sequence (including the proteolytic processing site: lys-arg) employed in the present study was a 255 nucleotide fragment shown in Figure 2.

This 255 nucleotide fragment was derived from plasmid pAO208, shown in Figure 3. The construction of plasmid pAO208 is described in detail below.

a. Construction of plasmid pAO208

The AOX1 transcription terminator was isolated from 20 μg of pPG2.0 [pPG2.0 = BamHI-HindIII fragment of pG4.0 (NRRL 15868) + pBR322] by StuI digestion followed by the addition of 0.2 μg SalI linkers (GGTCGACC). The plasmid was subsequently digested with HindIII and the

350 bp fragment isolated from a 10% acrylamide gel and subcloned into pUC18 (Boehringer Mannheim) digested with HindIII and SalI. The ligation mix was transformed into JM103 cells (that are widely available) and amp^R colonies were selected. The correct construction was verified by HindIII and SalI digestion, which yielded a 350 bp fragment, and was called pA0201.

5 μg of pA0201 was digested with HindIII, filled in using Klenow polymerase, and 0.1 μg of BglII linkers (GAGATCTC) were added. After digestion of the excess BglII linkers, the plasmid was reclosed and transformed into MC1061 cells. Amp^R cells were selected, DNA was prepared, and the correct plasmid was verified by BglII, SalI double digests, yielding a 350 bp fragment, and by a HindIII digest to show loss of HindIII site. This plasmid was called pA0202.

The alpha factor-GRF fusion was isolated as a 360 bp BamHI-PstI partial digest from pYSV201. Plasmid pYSV201 is the EcoRI-BamHI fragment of GRF-E-3 inserted into M13mp18 (New England Biolabs). Plasmid GRF-E-3 is described in EP 206,783. 20 μg of pYSV201 plasmid was digested with BamHI and partially digested with PstI. To this partial digest was added the following oligonucleotides:

5' AATTCGATGAGATTTCCTTCAATTTTTACTGCA 3'

3' GCTACTCTAAAGGAAGTTAAAAATG 5'.

Only the antisense strand of the oligonucleotide was kinase labelled so that the oligonucleotides did not polymerize at the 5'- end. After acrylamide gel electrophoresis (10%), the fragment of 385 bp was isolated by electroelution. This EcoRI- BamHI fragment of 385 bp was cloned into pA0202 which had been cut with EcoRI and BamHI. Routinely, 5 ng of vector cut with the appropriate enzymes and treated with calf intestine alkaline phosphatase, was ligated with 50 ng of the insert fragment. MC1061 cells were transformed, amp^r cells were selected, and DNA was prepared. In this case,

the resulting plasmid, pA0203, was cut with EcoRI and BglII to yield a fragment of greater than 700 bp. The α-factor-GRF fragment codes for the (1-40)leu^{27} version of GRF and contains the processing sites lys-arg-glu-ala-glu-ala.

The AOX1 promoter was isolated as a 1900 bp EcoRI fragment from 20 μg of pAOP3 and subcloned into EcoRI-digested pA0203. The development of pAOP3 is disclosed in EP 226,846 and described hereinbelow. MC1061 cells were transformed with the ligation reaction, amp^R colonies were selected, and DNA was prepared. The correct orientation contains a ≈376 bp HindIII fragment, whereas the wrong orientation has an ≈675 bp fragment. One such transformant was isolated and was called pA0204.

The parent vector for pA0208 is the HIS4, PARS2 plasmid pYJ32 (NRRL B-15891) which was modified to change the EcoRV site in the tet^R gene to a BglII site, by digesting PYJ32 with EcoRV and adding BglII linkers to create pYJ32(+BglII). This plasmid was digested with BglII and the 1.75 Kb BglII fragment from pA0204 containing the AOX1 promoter-α mating factor-GRF-AOX1 3' expression cassette was inserted. The resulting vector was called pA0208. An EcoRI digest of pA0208 yielded an 850 bp fragment + vector, while vector having the other orientation yielded a 1.1 Kb fragment + vector.

b. Construction of plasmid pAOP3:

1. Plasmid pPG2.5 [a pBR322 based plasmid containing the approximately 2.5 Kbp EcoRI-SalI fragment from plasmid pPG4.0, which plasmid contains the primary alcohol oxidase gene (AOX1) and regulatory regions and which is available in an E. coli host from the Northern Regional Research Center of the United States Department of Agriculture in Peoria, Illinois as NRRL B-15868] was linearized with BamHI.

2. The linearized plasmid was digested with BAL31;

3. The resulting DNA was treated with Klenow

fragment to enhance blunt ends, and ligated to EcoRI linkers;

4. The ligation products were transformed into E. coli strain MM294;

5. Transformants were screened by the colony hybridization technique using a synthetic oligonucleotide having the following sequence:

5'TTATTCGAAACGGGAATTCC.

This oligonucleotide contains the AOX1 promoter sequence up to, but not including, the ATG initiation codon, fused to the sequence of the EcoRI linker;

6. Positive clones were sequenced by the Maxam-Gilbert technique. All three positives had the following sequence:

5'...TTATTCGAAACGAGGAATTCC...3'.

They all retained the "A" of the ATG (underlined in the above sequence). It was decided that this A would probably not be detrimental; thus all subsequent clones are derivatives of these positive clones. These clones have been given the laboratory designation pAOP1, pAOP2 and pAOP3 respectively.

c. Construction of the expression vector pAO817

The hEGF gene and the AMF pre-pro sequence in the same translational direction were inserted into M13mp19 [New England Biolabs] by the following procedure:

10 μg of M13mp19 were digested with SmaI and EcoRI and the large, about 7240 bp plasmid fragment was isolated on a 0.8% agarose gel. The plasmid fragment and a 267 bp fragment contaning the AMF pre-pro sequence (including the proteolytic processing site: lys-arg) were ligated together by T4 DNA ligase. The 267 bp fragment containing the AMF pre-pro sequence was obtained by digesting 15mg of plasmid pAO208 with HindIII, filling in with Klenow-fragment DNA polynerase, and digesting with EcoRI. The digestion was run on a 1.7% agarose gel and the 267 bp fragment isolated.

The M13mp19-AMF pre-pro sequence ligation mixture was then transformed into JM103 cells and DNA from the plaque was characterized. Plasmid DNA was prepared from these cells and was digested with SalI, filled in the Klenow-fragment DNA polymerase, and cut with HindIII. The about 7400 bp plasmid fragment was isolated and ligated to the 160 bp hEGF gene fragment. The ligation mixture was transformed into JM101 cells and plaques were selected. Cells having the M13mp19 plasmid with both the EGF gene and the AMF pre-pro sequence in the same translational direction were called pEGF19-3.

In vitro mutagenesis was performed on pEGF19-3 to remove the polylinker of M13mp19 and to place the DNA sequence encoding AMF processing site: lys-arg, directly in front of the first codon of mature EGF. The mutagenesis was accomplished using standard techniques [Zoller and Smith, Meth. Enzymol. 100, 468 (1983)]. The mutagenizing oligonucleotide employed was of the following sequence:

5' TTC TTT GGA TAA AAG AAA TTC CGA TAG CGA GT 3'.

The screening oligonucleotide had the sequence: GATAAAAGAAATTCCGAT. The mutagenized plasmid was called pEGF19m-2.

EcoRI linkers of sequence GGAATTCC were added to the 3' end of the hEGF gene in the plasmid pEGF19m-2 by first digesting 20 μg of the plasmid with HindIII and then filling in with Klenow-fragment DNA polymerase. 1 μg of linkers and 20 μg of treated plasmid were ligated together and then digested with EcoRI to remove excess linkers. The ≈435bp EcoRI fragment was isolated on a 1.5% agarose gel. 15 μg of the plasmid pA0815 (the construction of which is described below) were digested with EcoRI and ligated to the 435 bp EcoRI fragment in a standard ligation reaction. The reaction was used to transform MC1061 cells and amp^R cells were selected. To determine which cells have a plasmid with the correct orientation of the AMF pre-pro sequence - hEGF gene

insert, plasmid DNA was prepared from the amp^R colonies and was digested with PstI. A correct construct yielded an about 1740 bp fragment. Colonies demonstrating the correct restriction pattern were called pA0816.

The complete, AOX1-promoter expression cassette was removed from pA0816 by digesting 15 μg of pA0816 with BglII and BamHI, and isolating the about 1670 bp fragment on a gel. The gel-purified fragment was then ligated to BamHI-cut pA0816. The ligation mix was used to transform MC1061 cells and amp^R colonies were selected. Colonies having plasmids comprised of two head-to-tail expression cassettes (referred to as pA0817) were identified by digestion with PstI, which gave fragments of 1827, 1497 and 9547 bp. The restriction map of pA0817 is shown in Figure 4.

d. Construction of plasmid pEGF819

Plasmid pEGF819 was constructed as follows: Plasmid pA0817 was digested with BglII and BamHI and the 3600 bp fragment containing two expression cassettes was isolated on a 0.8% agarose gel. 250 ng of fragment and 25 ng of BamHI-cut phosphatase-treated pA0817 were ligated together. The ligation was used to transform MC1061 cells and Amp^R cells were selected. DNA was prepared from the transformants and digested with BglII and BamHI. The plasmid was characterized by digesting with multiple restriction enzymes and comparing the resultant banding pattern with the banding pattern of other restriction enzyme digested vectors comprising known numbers of hEGF-sized expression cassettes. Based on the mobility of the hEGF-containing fragment of pEGF819 relative to the mobility of expression vectors comprising four and six expression cassettes, it was concluded that pEGF819 has five expression cassettes in tandem.

A restriction map of pEGF819 is shown in Figure 5.

e. Construction of plasmid pA0804

Plasmids pA0804 nad pA0807 are used in the construction of plasmid pA0815.

Plasmid pA0804 has been described in PCT Application No. WO 89/04320. Construction of this plasmid involved the following steps:

Plasmid pBR322 was modified as follows to eliminate the EcoRI site and insert a BglII site into the PvuII site:

pBR322 was digested with EcoRI, the protruding ends were filled in with Klenow Fragment of E. coli DNA polymerase I, and the resulting DNA was recircularized using T4 ligase. The recircularized DNA was used to transform E. coli MC1061 to ampicillin-resistance and transformants were screened for having a plasmid of about 4.37 kpb in size without an EcoRI site. One such transformant was selected and cultured to yield a plasmid, designated pBR322ΔRI, which is pBR322 with the EcoRI site replaced with the sequence:

5'-GAATTAATTC-3'

3'-CTTAATTAAG-5'.

pBR322ΔRI was digested with PvuII, and the linker having the sequence:

5'-CAGATCTG-3'

3'-GTCTAGAC-5'

was ligated to the resulting blunt ends employing T4 ligase. The resulting DNAs were recircularized, also with T4 ligase, and then digested with BglII and again recircularized using T4 ligase to eliminate multiple BglII sites due to ligation of more than one linker to the PvuII-cleaved pBR322ΔRI. The DNAs, treated to eliminate multiple BglII sites, were used to transform E. coli MC1061 to ampicillin-resistance. Transformants were screened for a plasmid of about 4.38 kbp with a BglII site. One such transformant was selected and cultured to yield a plasmid, designated pBR322ΔRIBGL, for further

work. Plasmid pBR322ΔRIBGL is the same as pBR322 RI except that pBR322ΔRIBGL has the sequence

5'-CAGCAGATCTGCTG-3'

3'-GTCGTCTAGACGAC-5'

in place of the PvuII site in pBR322ΔRI.

pBR322ΔRIBGL was digested with a SalI and BglII and the large fragment (approximately 2.97 kbp) was isolated. Plasmid pBSAGI5I, which is described in European Patent Application Publication No. 0 226 752, was digested completely with BglII and XhoI and an approximately 850 bp fragment from a region of the P. pastoris AOX1 locus downstream from the AOX1 gene transcription terminator (relative to the direction of transcription from the AOX1 promoter) was isolated. The BglII-XhoI fragment from pBSAGI5I and the approximately 2.97 kbp, SalI-BglII fragment from pBR322ΔRIBGL were combined and subjected to ligation with T4 ligase. The ligation mixture was used to transform E. coli MC1061 to ampicillin-resistance and transformants were screened for a plasmid of the expected size (approximately 3.8 kbp) with a BglII site. This plasmid was designated pA0801. The overhanging end of the SalI site from the pBR322ΔRIBGL fragment was ligated to the overhanging end of the XhoI site on the 850 bp pBSAGI5I fragment and, in the process, both the SalI site and the XhoI site in pA0801 were eliminated.

pBSAGI5I was then digested with ClaI and the approximately 2.0 kbp fragment was isolated. The 2.0 kbp fragment has an approximately 1.0-kbp segment which comprises the P. pastoris AOX1 promoter and transcription initiation site, an approximately 700 bp segment encoding the hepatitis B virus surface antigen ("HBsAg") and an approximately 300 bp segment which comprises the P. pastoris AOX1 gene polyadenylation signal and site-encoding segments and transcription terminator. The HBsAg coding segment of the 2.0 kbp fragment is terminated, at the end adjacent the 1.0 kbp segment with

the AOX1 promoter, with an EcoRI site and, at the end adjacent the 300 bp segment with the AOX1 transcription terminator, with a StuI site, and has its subsegment which codes for HBsAg oriented and positioned, with respect to the 1.0 kbp promoter-containing and 300 bp transcription terminator-containing segments, operatively for expression of the HBsAg upon transcription from the AOX1 promoter. The EcoRI site joining the promoter segment to the HBsAg coding segment occurs just upstream (with respect to the direction of transcription from the AOX1 promoter) from the translation initiation signal-encoding triplet of the AOX1 promoter.

For more details on the promoter and terminator segments of the 2.0 kbp, ClaI-site-terminated fragment of pBSAGI5I, see European Patent Application Publication No. 226,846 and Ellis et al., Mol. Cell Biol. 5, 1111 (1985).

Plasmid pA0801 was cut with ClaI and combined for ligation using T4 ligase with the approximately 2.0 kbp ClaI-site-terminated fragment from pBSAGI5I. The ligation mixture was used to transform E. coli MC1061 to ampicillin resistance, and transformants were screened for a plasmid of the expected size (approximately 5.8 kbp) which, on digestion with ClaI and BglII, yielded fragments of about 2.32 kbp (with the origin of replication and ampicillin-resistance gene from pBR322) and about 1.9 kbp, 1.48 kbp, and 100 bp. On digestion with BglII and EcoRI, the plasmid yielded an approximately 2.48 kbp fragment with the 300 bp terminator segment from the AOX1 gene and the HBsAg coding segment, a fragment of about 900 bp containing the segment from upstream of the AOX1 protein encoding segment of the AOX1 gene in the AOX1 locus, and a fragment of about 2.42 kbp containing the origin of replication and ampicillin resistance gene from pBR322 and an approximately 100 bp ClaI-BglII segment of the AOX1 locus (further upstream from the AOX1-encoding

segment than the first mentioned 900 bp EcoRI-BglII segment). Such a plasmid had the ClaI fragment from pBSAGI5I in the desired orientation, in the opposite undesired orientation, there would be EcoRI-BglII fragments of about 3.3 kbp, 2.38 kbp and 900 bp.

One of the transformants harboring the desired plasmid, designated pA0802, was selected for further work and was cultured to yield that plasmid. The desired orientation of the ClaI fragment from pBSAGI5I in pA0802 had the AOX1 gene in the AOX1 locus oriented correctly to lead to the correct integration into the P. pastoris genome at the AOX1 locus of linearized plasmid made by cutting at the BglII site at the terminus of the 800 bp fragment from downstream of the AOX1 gene in the AOX1 locus.

pA0802 was then treated to remove the HBsAg coding segment terminated with an EcoRI site and a StuI site. The plasmid was digested with StuI and a linker of sequence:

5'-GGAATTCC-3'

3'-CCTTAAGG-5'

was ligated to the blunt ends using T4 ligase. The mixture was then treated with EcoRI and again subjected to ligating using T4 ligase. The ligation mixture was then used to transform E. coli MC1061 to ampicillin resistance and transformants were screened for a plasmid of the expected size (5.1 kbp) with EcoRI-BglII fragments of about 1.78 kbp, 900 bp, and 2.42 kbp and BglII-ClaI fragment of about 100 bp, 2.32 kbp, 1.48 kbp, and 1.2 kbp. This plasmid was designated pA0803. A transformant with the desired plasmid was selected for further work and was cultured to yield pA0803.

Plasmid pA0804 was then made from pA0803 by inserting, into the BamHI site from pBR322 in pA0803, an approximately 2.75 kbp BglII fragment from the P. pastoris HIS4 gene. See, e.g., Cregg et al., Mol. Cell. Biol. 5, 3376 (1985) and European Patent Application

Publication Nos 180,899 and 188,677. pA0803 was digested with BamHI and combined with the HIS4 gene-containing BglII site-terminated fragment and the mixture subjected to ligation using T4 ligase. The ligation mixture was used to transform E. coli MC1061 to ampicillin-resistance and transformants were screened for a plasmid of the expected size (7.85 kbp), which is cut by SalI. One such transformant was selected for further work, and the plasmid it harbors was designated pA0804.

pA0804 has one SalI-ClaI fragment of about 1.5 kbp and another of abut 5.0 kbp and a ClaI-ClaI fragment of 1.3 kbp; this indicates that the direction of transcription of the HIS4 gene in the plasmid is the same as the direction of transcription of the ampicillin resistance gene and opposite the direction of transcription from the AOX1 promoter.

The orientation of the HIS4 gene in pA0804 is not critical to the function of the plasmid or of its derivatives with cDNA coding segments inserted at the EcoRI site between the AOX1 promoter and terminator segments. Thus, a plasmid with the HIS4 gene in the orientation opposite that of the HIS4 gene in pA0804 would also be effective for use in accordance with the present invention.

f. Construction of plasmid pA0807

1. Preparation of f1-ori DNA

f1 bacteriophage DNA (50 μg) was digested with 50 units of Rsa I and Dra I (according to manufacturer's directions) to release the ≈458 bp DNA fragment containing the f1 origin of replication (ori). The digestion mixture was extracted with an equal volume of phenol: chloroform (V/V) followed by extracting the aqueous layer with an equal volume of chloroform and finally the DNA in the aqueous phase was precipitated by adjusting the NaCl concentration to 0.2M and adding 2.5 volumes of absolute ethanol. The mixture was allowed to stand on ice (4°C) for 10 minutes and the DNA precipitate

was collected by centrifugation for 30 minutes at 10,000 x g in a microfuge at 4°C.

The DNA pellet was washed 2 times with 70% aqueous ethanol. The washed pellet was vacuum dried and dissolved in 25 µl of TE buffer. This DNA was electrophoresed on 1.5% agarose gel and the gel portion containing the ≈458 bp f1-ori fragment was excised out and the DNA in the gel was electroeluted onto DE81 (Watman) paper and eluted from the paper in 1M NaCl. The DNA solution was precipitated as detailed above and the DNA precipitate was dissolved in 25 µl of TE buffer (f1-ori fragment).

2. Cloning of f1-ori into Dra I sites of pBR322

pBR322 (2 µg) was partially digested with 2 units Dra I (according to manufacturer's instructions). The reaction was terminated by phenol:chloroform extraction followed by precipitation of DNA as detailed in step 1 above. The DNA pellet was dissolved in 20 µl of TE buffer. About 100 ng of this DNA was ligated with 100 ng of f1-ori fragment (step 1) in 20 µl of ligation buffer by incubating at 14°C for overnight with 1 unit of T4 DNA ligase. The ligation was terminated by heating to 70°C for 10 minutes and then used to transform E. coli strain JM103. Amp^R transformants were pooled and superinfected with helper phage R408. Single stranded phage were isolated from the media and used to reinfect JM103. Amp^R transformants contained pBRf1-ori which contains f1-ori cloned into the Dra I sites (nucleotide positions 3232 and 3251) of pBR322.

3. Construction of plasmid pA0807

pBRf1-ori (10 µg) was digested for 4 hours at 37°C with 10 units each of Pst I and Nde I. The digested DNA was phenol:chloroform extracted, precipitated and dissolved in 25 µl of TE buffer as detailed in step 1 above. This material was electrophoresed on a 1.2% agarose gel and the Nde I - Pst

I fragment (approximately 0.8 kb) containing the f1-ori was isolated and dissolved in 20 μl of TE buffer as detailed in step 1 above. About 100 ng of this DNA was mixed with 100 ng of pA0804 that had been digested with Pst I and Nde I and phosphatase-treated. This mixture was ligated in 20 μl of ligation buffer by incubating for overnight at 14°C with 1 unit of T4 DNA ligase. The ligation reaction was terminated by heating at 70°C for 10 minutes. This DNA was used to transform E. coli strain JM103 to obtain pA0807.

g. Construction of plasmid pA0815

Plasmid pA0815 was constructed by mutagenizing plasmid pA0807 to change the ClaI site downstream of the AOX1 transcription terminator in pA0807 to a BamHI site. The oligonucleotide used for mutagenizing pA0807 had the following sequence:

5' GAC GTT CGT TTG TGC GGA TCC AAT GCG GTA GTT TAT 3'.

The mutagenized plasmid was called pA0807-Bam. Plasmid pA0804 was digested with BglII and 25 ng of the 2400 bp fragment were ligated to 250 ng of the 5400 bp BglII fragment from BglII-digested pA0807-Bam. The ligation mix was transformed into MC1061 cells and the correct construct was verified by digestion with Pst/BamHI to identify 6100 and 2100 bp sized bands. The correct construct was called pA0815.

Example 2

Development of hEGF-secreting strains

1. Mut⁻ strains

20 μg of the expression vector pA0817 were digested with BglII, which releases the AOX1-ended tandem expression cassette. The linear DNA fragment obtained by digestion (5 μg) was transformed into the P. pastoris strain GS115 (ATCC 20864) by the spheroplast method [Cregg et al., Mol. Cell. Biol. 5, 3376 (1985)]. His⁺ cells were selected and the methanol utilization phenotype (Mut) of the cells was determined as follows:

His$^+$ transformants were plated on minimal glucose (2%) master plates to obtain colonies originating from single cells. After overnight incubation at 30°C, the masters were replica-plated to minimal glucose plates and plates containing no carbon source to which methanol was added in vapor phase. This is accomplished by adding an aliquot, approximately 200 µl, of methanol to the underside of the top of a covered petri dish. The plates were incubated at 30°C for 4-6 days with additional MeOH added in the vapor phase every two days. Colonies showing visible growth were scored as Mut$^+$ and those with no visible growth were scored as Mut$^-$.

Approximately 15% of the cells were His$^+$Mut$^-$, indicating that the expression vector integrated correctly at the AOX1 locus and disrupted the AOX1 gene. Southern analysis of an EcoRI digest of the transformants, using the plasmid pA0803 as probe, confirmed the disruption of the AOX1 gene and showed the number of expression units integrated. The strains were named as follows:

Name	Phenotype	Site of Integration	Copy Number
G-EGF817S10	Mut$^-$His$^+$	AOX1	One
G-EGF817S7	Mut$^-$His$^+$	AOX1	One
G-EGF817S9	Mut$^-$His$^+$	AOX1	Multiple

In the above table copy number refers to the number of BglII fragments integrated. Each BglII fragment is comprised of two EGF expression cassettes.

2. Mut$^+$ strains

P. pastoris strain GS115 (ATCC 20864) was transformed with 5 µg of uncut vector pA0817 using the spheroplast method of transformation. In this type of transformation the plasmid will integrate by addition into the P. pastoris genome at a site of homology between the plasmid and the host strain. The transformants were

screened for His$^+$Mut$^+$ phenotype, and several were picked for Southern analysis. An EcoRI digest was probed with plasmid pYM4 [pYM4 was obtained by digesting pYM30 (NRRL B-15890) with ClaI and religating the ends] and the hybridization pattern revealed two of the six had appropriate integrations:

Name	Phenotype	Site of Integration	Copy Number
G+EGF817S1	Mut$^+$His$^+$	HIS4	One
G+EGF817S6	Mut$^+$His$^+$	HIS4	One

P. pastoris strain GS115 was transformed with 1 μg of uncut plasmid pEGF819 using the spheroplast method of transformation. The transformants were screened for His$^+$ Mut$^+$ phenotype and several were picked for Southern analysis as described for the Mut$^+$ pAO817 transformants. Isolate G+EGF819S4 was further characterized by isolating a major portion of plasmid pEGF819 from the genome of G+EGF819S4, then digesting with multiple restriction enzymes, as follows:

Genomic DNA was isolated from a culture of G+EGF819S4 grown from a single colony. The DNA was digested with the restriction enzyme BglII and separated by agarose gel electrophoresis. As a size marker, plasmid pEGF819 was also digested with BglII and electrophoresed on the same gel. Previous Southern analysis of BglII-digested G+EGF819S4 DNA probed with pBR322 containing AOX1 5' and 3' regions, or HIS4, or an EGF-specific oligonucleotide, indicated that the BglII genomic fragment of G+EGF819S4 containing multiple expression cassettes is approximately the same size as the BglII fragment from the plasmid. Therefore, the area of the genomic digest corresponding in size to the BglII fragment from the plasmid was eluted from the gel and cloned into pUC19 to create plasmid pEGF772-3.

Plasmids pEGF772-3 and pEGF819 were analyzed by digestion with several restriction enzymes including

EcoRI, HindIII, BamHI, EcoRV, SalI, SacI, PstI, XbaI, and double digests with EcoRI/BamHI, HindIII/BamHI. These digests indicated that although the genomic BglII fragment of G+EGF819S4 and the BglII fragment from pEGF819 are the same size, the 5' ends of the fragments are different. The 5' end of the pEGF772-3 insert was sequenced to precisely identify how the two fragments differed.

Partial sequence analysis of pEGF772-3 revealed that the 5' end of the genomic BglII fragment contains the 3' portion of the AOX1 gene instead of the expected AOX1 promoter region from the first cassette. A subclone of the genomic fragment was generated in order to sequence the 5' end of the fragment completely and to more exactly determine the site of integration of the expression plasmid. Thus, pEGF772-3 was digested with EcoRI and reclosed with T4 DNA ligase. This procedure eliminated the majority of the genomic fragment leaving approximately 2500 bp of the 5' end. The new plasmid was called pEGF772-8.

The sequence of this 5' region (approximately 2200 bp sequenced) of the genomic BglII fragment indicated that the 3' end of the AOX1 gene including its transcription termination region are intact. It appears that expression plasmid pEGF819 integrated into the AOX1 locus by recombination of the transcription termination region of the first expression cassette with the homologous region in the AOX1 gene. This recombination resulted in the loss of the first cassette as well as the 2400 bp of pBR322 found upstream of the first expression cassette in the plasmid.

The 3' end of the AOX1 gene including its transcription termination segment contains approximately 1500 bp of DNA, which is approximately the same size as one of the EGF expression cassettes. Thus, although the genomic BglII fragment of G+EGF819S4 is approximately the same size as BglII-linearized pEGF819, which contains

five copies of the EGF cassette, the expression strain contains only four copies of the expression cassette, as well as the 3' end of AOX1.

To further characterize the expression cassettes integrated in G+EGF819S4, plasmid pEGF772-3 was digested with EcoRI which liberates four individual AMF-EGF fusion coding regions. The band corresponding to the cassettes was isolated from vector bands from an agarose gel and cloned into M13mp18. Twelve of these M13 clones were sequenced by the Sanger dideoxy method, and all were found to have the expected nucleic acid sequence.

Example 3

Fermentation of EGF strains

a. Fermentor start-up and general operation

The 2-liter fermentors (L.H. Fermentation, Hayward, CA; Biolafitte, LSL Biolafitte, Princeton, NJ) were autoclaved at a 700 ml volume containing 225 ml of 10X basal salts (52 ml/l 85% phosphoric acid, 1.8 g/l Calcium Sulphate-$2H_2O$, 28.6 g/l Potassium Sulfate, 23.4 g/l Magnesium Sulfate-7H2O, 6.5 g/l Potassium Hydroxide) and 30 g glycerol. After sterilization, 3 ml of a YTM_4 trace salts solution (5.0 ml/l Sulfuric Acid, 65.0 g/l Ferrous Sulfate-$7H_2O$, 6.0 g/l Copper Sulfate-$5H_2O$, 20.0 g/l Zinc Sulfate-$7H_2O$, 3.0 g/l Manganese Sulfate-H_2O, 0.1 g/l Biotin) was added and the pH adjusted to 5.0 with the addition of concentrated Ammonium Hydroxide; the pH was then controlled at 5.0 with the addition of a 20% Ammonium Hydroxide solution containing 0.1% Struktol J673 antifoam (Struktol Co., Stow, OH) throughout the fermentation. Excessive foaming was controlled throughout the fermentation by addition of Struktol J693 antifoam when foam contacted a foam sensor in the fermentor. The fermentors were then inoculated with a 10-50 ml volume of inoculum (overnight shake flask culture in phosphate-buffered 0.65% Yeast Nitrogen Base, pH6, containing 2% glycerol). Upon exhaustion of the

initial glycercl charge, a glycerol feed was started as described below. The dissolved oxygen of the fermentation was maintained above 20% of air saturation by increasing the air flow rate up to 3 liter/minute and agitation speed up to 1500 rpm during the fermentation.

Ten-liter fermentations (in a 15-liter Biolafitte fermentor) were started in a 7.0 liter volume containing 4 liters of 10X basal salts and 520 g of glycerol for the Mut^+ methanol fed-batch protocol. After sterilization, 30 ml each of YTM_4 and IM_1 trace salts solutions were added and the pH was adjusted and subsequently controlled at 5.0 with the addition of ammonia gas throughout the fermentation. Excessive foaming was controlled with the addition of 5% Struktol J673 antifoam. The fermentor was inoculated with a volume of 200-500 ml. Upon exhaustion of the initial glycerol charge, a feed was started as outlined below. The dissolved oxygen was maintained above 20% by increasing the air flow rate up to 40 liter/minute, the agitation up to 1000 rpm and/or the pressure of the fermentor up to 1.5 bar during the fermentation.

b. Growth of Mut^- strains in one-liter fermentors

(1) Mut^- (NL) mixed-feed fed batch fermentation

Run 413:G-EGF817S10
Run 419:G-EGF817S9
Run 422:G-EGF817S9
Run 423:G-EGF817S10
Run 434:G-EGF817S9

After the glycerol batch phase was completed, a 50% (by weight) glycerol feed, containing 12 ml/l YTM_4 trace salts was started at 5.4 ml/h for the 2-liter fermentor. After 6 hours of glycerol feeding, the glycerol feed was decreased to 3.6 ml/h (36 ml/h at 10-liters) and a methanol feed containing 12 ml/l YTM4 trace salts was initiated at 1.1 ml/h for the 2-liter

fermentor. After 5 hours, the methanol feed was adjusted to give a residual methanol concentration of up to about 1%, preferably between 0.2 and 0.8%. The fermentation was sampled periodically and harvested 36-50 hr after the methanol feed was initiated.

(2) Mut⁻ methanol-fed-batch

Run 425:G-EGF817S9

Run 426:G-EGF817S10

After the glycerol batch phase was completed, an induced fed-batch phase was initiated by adding methanol to the fermentor to maintain a residual methanol concentration between 0.2 and 0.8%. The fermentor was sampled periodically and harvested after 167 hr growth on methanol.

(3) Alternative procedure for production of 1-52 hEGF

Run 470:G-EGF817S9

A two liter LH fermentor containing 400 ml 10X basal salts, 80 g glycerol, and deionized water (to 1 liter) was sterilized. After sterilization and cooling, 3 ml YTM_4 + biotin solution was added and 20% NH_4OH used to bring pH to 3.6. The fermentor was inoculated with 60 ml of inoculum of Mut⁻ cells and the pH controller set at 5.0. During batch growth, the agitation speed was adjusted upward periodically to maintain a dissolved oxygen tension above 20% air saturation. After exhaustion of the initial glycerol charge, a 50% solution of glycerol containing 12 ml/l YTM_4 + biotin was pumped into the fermentor at the rate of 20 ml/h. Four and one-half hours later, the glycerol feed rate was decreased to 10 ml/hr and a feed of methanol containing 12 ml/l YTM^4 + biotin was started at 1.0 ml/h. Three hours later the methanol feed rate was doubled. After ninety minutes at 2 ml/h, the methanol feed rate was adjusted to 3.8 ml/h and maintained constant until harvest at 13.5 hours after the methanol feed was first initiated.

c. Growth of Mut$^+$ strains in two-liter and 14-liter fermentors

(1) Mut$^+$ methanol-fed-batch

Run 483:G+EGF819S4 (2L)

Run 464:G+EGF817S1 (2L)

Run 490:G+EGF819S4 (14L)

After glycerol exhaustion, a 50% glycerol feed, containing 12 ml/l YTM$_4$ trace salts, was started at 12 ml/h for the 2-liter or 200 ml/h for the 10-liter fermentor and run for a total of 7 hours. After 6 hours on the glycerol feed, the methanol feed, containing 12 ml/l YTM$_4$ trace salts, was started at 1.1 ml/h for the 2-liter and 11 ml/h for the 10-liter fermentor for 5 minutes. When a rise in dissolved oxygen was seen after the methanol feed was shut-off, the methanol feed was turned back on for another 5 minute interval. The latter process was repeated several times until an immediate response in the dissolved oxygen was observed to the methanol feed cessation; once this occurred, the methanol feed was increased by 20% per hour at 30 minute intervals. The methanol feed was increased until a feed rate of 7.6 ml/h for the 2-liter or 90 ml/h for the 10-liter fermentor was reached. The fermentation was then carried out for 40-60 hours for the 2-liter or 25-35 hours for the 10-liter fermentor.

(2) Alternative procedure for growth of Mut$^+$ strains

Fifteen-liter fermentations employing strain G+EGF819S4 (in a 15-liter Biolafitte fermentor) were started in a six-liter volume containing four liters of 10X basal salts and 400 g of glycerol. After sterilization, 25 ml of PTM$_1$ trace salts solution [6.0 g/L cupric sulfate$\cdot 5H_2O$, 0.08 g/L sodium iodide, 3.0 g/L manganese sulfate$\cdot H_2O$, 0.2 g/L sodium molybdate$\cdot 2H_2O$, 0.02 g/L boric acid, 0.05 g/L cobalt chloride, 56.0 g/L ferrous sulfate$\cdot 7H_2O$, 0.2 g/L biotin and 5.0 ml/L sulfuric acid (conc)] were added and the pH was adjusted and

subsequently controlled at 5.0 with the addition of ammonia gas throughout the fermentation. Excessive foaming was controlled with the addition of 5% Struktol J673 antifoam. The fermentor was inoculated with a volume of 500 ml of an overnight culture (OD_{600}= 1 to 4) in YNB, 2% glycerol, 0.1 M phosphate, pH 6. The dissolved oxygen was maintained above 20% by increasing the air flow rate up to 20 liter/minute, the agitation up to 1000 rpm and/or the pressure of the fermentor up to 1.5 bar during the fermentation.

After exhaustion of the initial glycerol charge, a 50% glycerol feed, containing 12 ml/L PTM_1 trace salts, was initiated at a rate of 120 ml/h; the glycerol feed continued for 6 hours, at which time the methanol feed, containing 12 ml/L PTM_1 trace salts, was started at a rate of 20 ml/h. The methanol feed was increased by 20% each hour at half hour intervals until a methanol feed rate of 100 ml/h was reached. The fermentation was then continued for 25-35 hours.

The conditions for 2- and 250-liter fermentations were scaled proportionately from the 15-liter fermentation, except that the final methanol feed rate was limited to the highest rate at which the dissolved oxygen concentration could be maintained above 20% air saturation. In 2- and 250-liter fermentations, the pH was controlled with NH_4OH rather than NH_3, and in the 250 liter fermentation, the air sparge was supplemented with O_2in some runs.

Example 4

Results of fermentations

a. Mut$^-$ strains

The time course of cell growth and EGF expression in four fermentor runs of two Mut$^-$ strains was investigated. Cell growth for strains G-EGF817S9 and G-EGF817S10 under a methanol fed batch protocol was similar, yielding about 300 g/l wet cells after 167 h.

However, the multicopy strain G-EGF817S9 produced 400 mg/l of EGF, twice as much as the strain with only two copies of the EGF expression cassette. The maximum concentration of EGF was reached after about 120 hours growth on methanol.

A similar pattern is observed for the two strains growing more rapidly under the mixed-feed protocol. In this protocol, both strains again grew up to more than 300 g/l, and the 400 mg/l of EGF produced by the multicopy strain is again higher than that produced by the double copy strain.

The two fermentation protocols, i.e., the methanol fed batch and mixed feed protocols, were then carried out with the multicopy strain G-EGF817S9. The results demonstrate that a dramatically reduced time on methanol is required to produce EGF using mixed feed compared to using methanol alone, 35 hr vs. 120 hr, respectively. The initial batch growth on glycerol to build up cell mass adds another 24 h to the overall process time. The EGF productivities for the methanol and mixed feed modes are 3 $mgl^{-1}h^{-1}$ and 7 $mgl^{-1}h^{-1}$, respectively.

b. Mut$^+$ strains

The time course of hEGF production at both 1L (Run 483) and 10L (Run 490) volumes in fermentations employing the Mut$^+$ strain G+EGF819S4 was investigated. Results are summarized in Table I.

Table I

Time Course of hEGF Production from Strain G+EGF819S4

Reaction scale, l	Time on methanol feed, hr	Approximate EGF concn in broth, mg/l
1	4	35
	7	75
	12	130
	20	305
	44	465
10	0	25
	3	60
	5	130
	7	160
	11	260
	13	340
	15	330
	18	345
	20	290
	26	325
	33	400
	40	610

The higher hEGF production seen in these fermentations, 500-600 mg/L, as compared to the Mut$^-$ fermentations is due to the higher copy number of G+EGF819S4 (4) rather than the Mut$^+$ phenotype. A Mut$^+$ strain carrying two copies of the EGF gene, G+EGF817S1, produced hEGF at concentrations similar to those seen in a Mut$^-$ strain carrying two copies of the hEGF gene.

Example 5

Analysis of secreted EGF

1. Western analysis

The first mode of analysis for evaluation of Pichia-produced EGF was the Western blot. Because antisera against hEGF can have low cross-reactivity to mEGF, it was necessary to obtain human EGF standard and antisera, instead of mEGF and anti-mEGF, respectively, for our analyses. The human EGF reagents were acquired commercially from Amgen (standard) and Biomedical

Technologies Inc., Stoughton, MA (antisera). Electrophoresis of broth samples was conducted on a 15% polyacrylamide gel. Western blot analysis revealed that almost all the immunoreactive material was found in a single band which was of the same size as the human EGF standard. In several of the samples, (Runs 422, 423, 425), a larger molecular weight species, approximately 33 KD in size, was also seen to react with the antisera. By this analysis, the amount of EGF produced in runs with the multicopy strain appeared to be about twice as much as that produced in runs with the two copy strain.

2. Stained gels

Protein bands on the acrylamide gels were also visualized by staining with Coomasie blue. The primary protein species in the gel typically has an electrophoretic mobility similar to that of standard EGF. A further confirmation of the relative abundance of EGF protein in the broth was given by total protein assay of the broth. In the sample from Run 423, total TCA precipitable protein determined by the Lowry assay (100 mg/l ± 10 mg/l) was on the low end of the EGF concentration range estimated by Western blot with f-met-EGF standard (100-180 mg/l).

3. Separation of EGF peptides on HPLC

Three peptides that eluted separately on reverse phase HPLC were purified to homogeneity by analytical HPLC. These peptides are designated with the numbers 1, 2, and 4 in the order of decreasing elution time. Approximately 50 μg of each peptide was obtained in a volatile buffer. Peaks 1 and 2 were purified from Run 470; Peak 4 was purified from Run 425. Peak 3 was not purified due to its relatively low concentration.

All three peaks were submitted to quantitative amino acid analysis after hydrolysis in 6N HCl containing 0.1% phenol. The compositions of hEGF Peaks 1 and 2 are consistent with an hEGF peptide that lacks a single arginine at the C-terminus. The composition of Peak 4,

on the other hand, shows lower amounts of leucine and glutamic acid, and suggests decreased yields of lysine and aspartic acid. All three peptides, however, possess an authentic amino-terminus as determined by automated Edman degradation. This suggests that the difference in the composition of Peptide 4 results from alteration at its carboxy-terminus.

In an effort to determine the carboxy-terminal sequence of the peptides, they were each digested with carboxypeptidase Y (CPY) and the amino acids released over time were measured on an amino acid analyzer. The most rapidly released amino acids were leucine followed by glutamic acid. Thus, it was concluded that Peak 1 is a 1-52 product of the originally translated peptide; the carboxy-terminal arginine was probably removed by proteolysis during fermentation. Peak 2 gave similar results as Peak 1, but Peak 4 did not yield any amino acids. This negative result was difficult to interpret, but could have been the result of a carboxy-terminal residue that is difficult for CPY to release, such as a lysine which occurs at position 48 of hEGF.

To determine if the tryptophan residues at positions 49 and 50 were absent in Peak 4, one microgram of each peptide (1,2,4) was submitted to reverse phase HPLC on a chromatography system equipped with a diode array detector (Hewlett Packard 1090). Absorbance at 280 nm and 210 nm was collected simultaneously for each peptide and the ratio 210 nm/280 nm was calculated both on the basis of peak height and integrated area. This ratio should be indicative of the tryptophan and tyrosine content of a peptide. More specifically, the ratio reflects the relative number of peptide bonds (contributors to 210 nm absorbance) to the number of tryptophan residues (contributors to 280 nm absorbance). Tryptophan residues, when present in a sequence, tend to mask the smaller contribution of tyrosine to the absorbance at 280 nm.

The 210 nm/280 nm absorbance ratios of hEGF Peaks 1 and 2 were in the same range as those of other proteins that have a similar content of tryptophan. Peak 4, however, had a larger absorbance ratio which indicated the absence of any tryptophan residues. The unusually high number of tyrosine residues that remain in Peptide 4 depress the value of the absorbance ratio slightly.

The data from total amino acid analysis, N-terminal sequence, carboxypeptidase Y digestion, and UV absorbance ratios indicated that both Peaks 1 and 2 are 1-52 forms of hEGF while Peak 4 was a considerably shorter form.

The molecular weight of peptide 4 was subsequently determined by mass spectrometry and was consistent with an hEGF peptide comprised of residues 1-48. Carboxy peptidase digests of the peptide confirmed that the C-terminal peptide is the 48th residue, lysine.

4. Amino acid sequencing

Fractions containing the HPLC peaks at 22.47 min, 28.74 min, and 31.44 min were collected, and eight residues were sequenced on an automated gas phase protein microsequencer. Both the 22.47 min and 31.44 min peaks yielded the correct N-terminal sequence for EGF for the first eight residues. The peak at 28.74 min was not related to EGF.

5. Stability of secreted EGF in fermentation broth of Pichia pastoris

HPLC analysis of hEGF in the broth during the time course of the fermentation runs revealed that the 1-48 peptide was much more stable than the longer forms. The longer forms could be seen early after induction during the run. After 24h growth on methanol, peptide 4 would accumulate, apparently as a degradation product of the other forms. Peptide 4 was very stable under fermentation conditions, persisting and accumulating for up to six days in the longer fermentation protocols.

This unexpected high stability makes production and purification of this form of hEGF much simpler than that of the longer forms.

6. Biological activity

The 1-48 hEGF peptide was tested for biological activity both in in vitro cell mitogenic assays and in vivo in stimulation of gastric ulcer healing. The peptide was observed to have high biological activity in both types of tests.

The invention has been described in detail with reference to particular embodiments thereof. It will be understood, however, that variations and modifications can be effected within the spirit and scope of the invention.

CLAIMS:

1. A DNA fragment comprising an expression cassette, wherein said expression cassette comprises, in the direction of transcription, the following DNA sequences:

(i) a promoter region of a methanol responsive gene of a methylotrophic yeast,

(ii) a DNA sequence encoding a polypeptide consisting of:

(a) the S. cerevisiae AMF pre-pro sequence, including the processing site: lys-arg, and

(b) a DNA sequence encoding an EGF peptide; and

(iii) a transcription terminator functional in a methylotrophic yeast,

wherein said DNA sequences are operationally associated with one another for transcription of the sequences encoding said polypeptide.

2. A DNA fragment according to Claim 1 further comprising at least one selectable marker gene and a bacterial origin of replication.

3. A DNA fragment according to Claim 2 wherein said fragment is contained within a circular plasmid.

4. A DNA fragment according to Claim 1 wherein said sequence encoding an EGF peptide encodes the 1-53 or 1-48 form of EGF.

5. A DNA fragment according to Claim 1 wherein said methylotrophic yeast is a strain of Pichia pastoris.

6. A DNA fragment according to Claim 5 wherein said methanol responsive gene of a methylotrophic yeast and the transcription terminator are both derived from the P. pastoris AOX1 gene.

7. A DNA fragment according to Claim 6 further comprising 3'- and 5'-ends having sufficient homology with a target gene of a yeast host for said DNA fragment to effect site directed integration of said fragment into said target gene.

8. A DNA fragment according to Claim 1 further comprising 3'- and 5'-ends having sufficient homology with a target gene of a yeast host for said DNA fragment to effect site directed integration of said fragment into said target gene.

9. A DNA fragment according to Claim 1 containing multiple copies of said expression cassette.

10. A DNA fragment according to Claim 9 wherein said multiple copies of said expression cassette are oriented in head-to-tail orientation.

11. A DNA fragment according to Claim 7, which is derived from a BglII digest of the Pichia expression vector pAO817.

12. A DNA fragment according to Claim 7, which is the Pichia expression vector pEGF819.

13. A DNA fragment according to Claim 7, which is derived from a BglII-BamHI digest of the Pichia expression vector pA0816.

14. A methylotrophic yeast cell transformed with the DNA fragment of Claim 1.

15. A methylotrophic yeast cell according to claim 14 wherein said yeast is a strain of Pichia pastoris.

16. A methylotrophic yeast cell transformed with the DNA fragment of Claim 4.

17. A methylotrophic yeast cell according to Claim 16 wherein said yeast is a strain of Pichia pastoris.

18. A P. pastoris cell transformed with the DNA fragment of Claim 5.

19. A P. pastoris cell transformed with the DNA fragment of Claim 6.

20. A P. pastoris cell transformed with the DNA fragment of Claim 7.

21. A P. pastoris cell according to Claim 20, wherein said cell is selected from strain G-EGF817S10, G-EGF817S7, G-EGF817S9, G+EGF817S1, G+EGF817S4 or G+EGF817S6.

22. A methylotrophic yeast cell transformed with the DNA fragment of Claim 8.

23. A methylotrophic yeast cell transformed with the DNA fragment of Claim 9.

24. A methylotrophic yeast cell transformed with the DNA fragment of Claim 10.

25. A P. pastoris cell transformed with the DNA fragment of Claim 11.

26. A P. pastoris cell according to Claim 25 wherein said cell is selected from strain G-EGF817S10, G-EGF817S7 or G-EGF817S9.

27. A P. pastoris cell transformed with the DNA fragment of Claim 12.

28. A P. pastoris cell according to Claim 27 wherein said cell is selected from strain G+EGF817S1, G+EGF817S4 or G+EGF817S6.

29. A culture of viable P. pastoris cells according to Claim 13.

30. A culture of viable P. pastoris cells according to Claim 21.

31. A process for producing EGF, said process comprising growing the cells of Claim 14 under conditions allowing the expression of said expression cassette(s) in said cells, and the secretion of said EGF product into the culture medium.

32. A process according to Claim 31 wherein said methylotrophic yeast is a strain of Pichia pastoris.

33. A process according to Claim 31 wherein said cells are grown in a medium containing methanol as a carbon source.

34. A process according to Claim 31 wherein said cells have the Mut$^-$ phenotype.

35. A process according to Claim 31 wherein said cells have the Mut$^+$ phenotype.

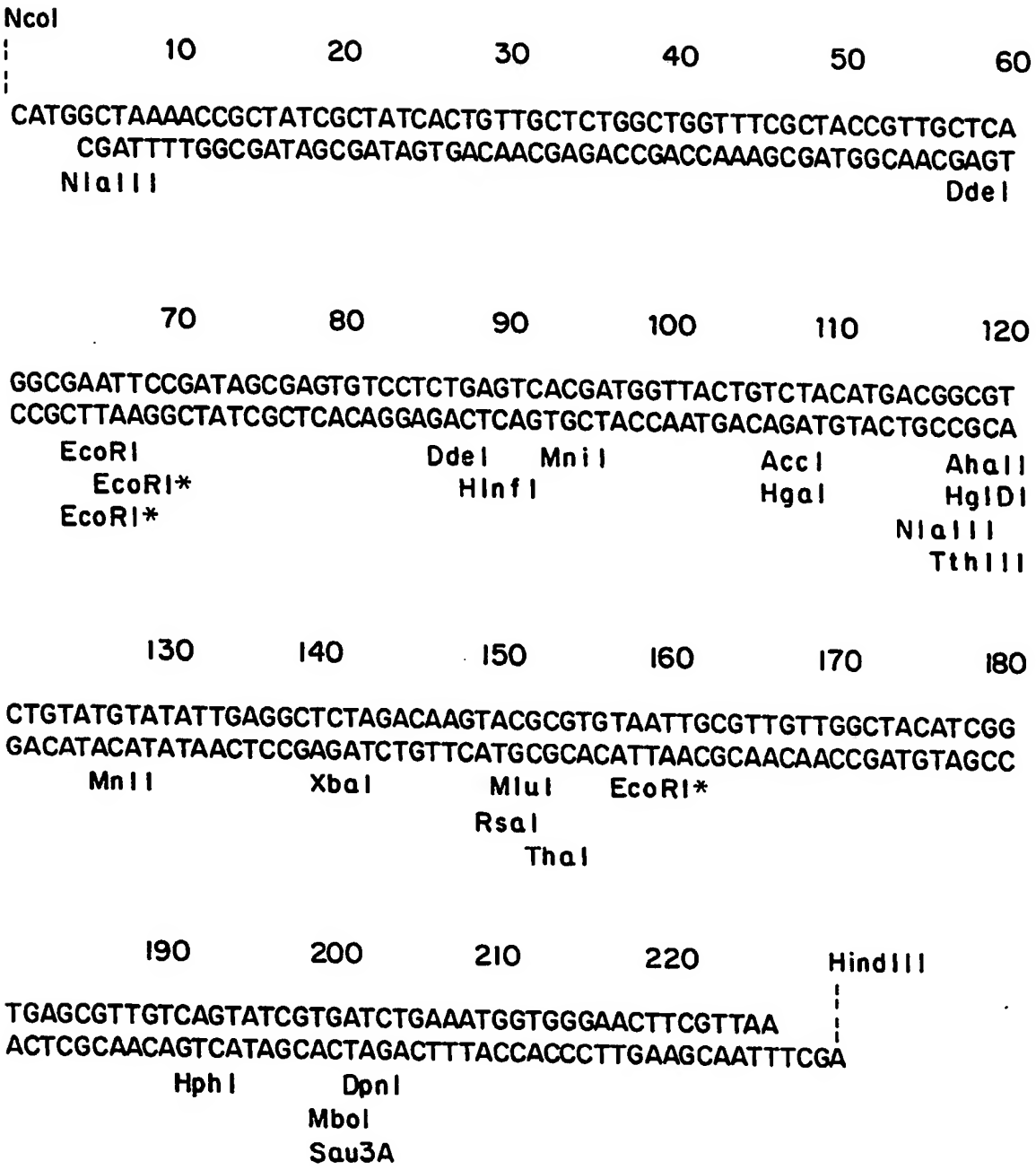
NcoI
10 20 30 40 50 60
CATGGCTAAAACCGCTATCGCTATCACTGTTGCTCTGGCTGGTTTCGCTACCGTTGCTCA
CGATTTTGGCGATAGCGATAGTGACAACGAGACCGACCAAAGCGATGGCAACGAGT
NlaIII
DdeI
70 80 90 100 110 120
GGCGAATTCCGATAGCGAGTGTCCTCTGAGTCACGATGGTTACTGTCTACATGACGGCGT
CCGCTTAAGGCTATCGCTCACAGGAGACTCAGTGCTACCAATGACAGATGTACTGCCGCA
EcoRI
EcoRI*
EcoRI*
DdeI MnlI
HinfI
AccI
HgaI
AhaII
HgiDI
NlaIII
TthIII
130 140 150 160 170 180
CTGTATGTATATTGAGGCTCTAGACAAGTACGCGTGTAATTGCGTTGTTGGCTACATCGG
GACATACATATAACTCCGAGATCTGTTCATGCGCACATTAACGCAACAACCGATGTAGCC
MnlI
XbaI
MluI
RsaI
TaiI
EcoRI*
190 200 210 220 HindIII
TGAGCGTTGTCAGTATCGTGATCTGAAATGGTGGGAACTTCGTTAA
ACTCGCAACAGTCATAGCACTAGACTTTACCACCCTTGAAGCAATTTCGA
HphI
DpnI
MboI
Sau3A

FIG. 1

LEADER SEQUENCE OF S. CEREVISIAL ALPHA MATING FACTOR PRECURSOR
(268bp)

```
        10                  30                  50
         .         .         .         .         .         .
ATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGCATTAGCTGCT
MetArgPheProSerIlePheThrAlaValLeuPheAlaAlaSerSerAlaLeuAlaAla

        70                  90                  110
         .         .         .         .         .         .
CCAGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGGT
ProValAsnThrThrThrGluAspGluThrAlaGlnIleProAlaGluAlaValIleGly

        130                 150                 170
         .         .         .         .         .         .
TACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACAAAT
TyrSerAspLeuGluGlyAspPheAspValAlaValLeuProPheSerAsnSerThrAsn

        190                 210                 230
         .         .         .         .         .         .
AACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAGGGGTA
AsnGlyLeuLeuPheIleAsnThrThrIleAlaSerIleAlaAlaLysGluGluGlyVal

        250
         .
TCTTTGGATAAAAGA
SerLeuAspLysArg
```

FIG. 2

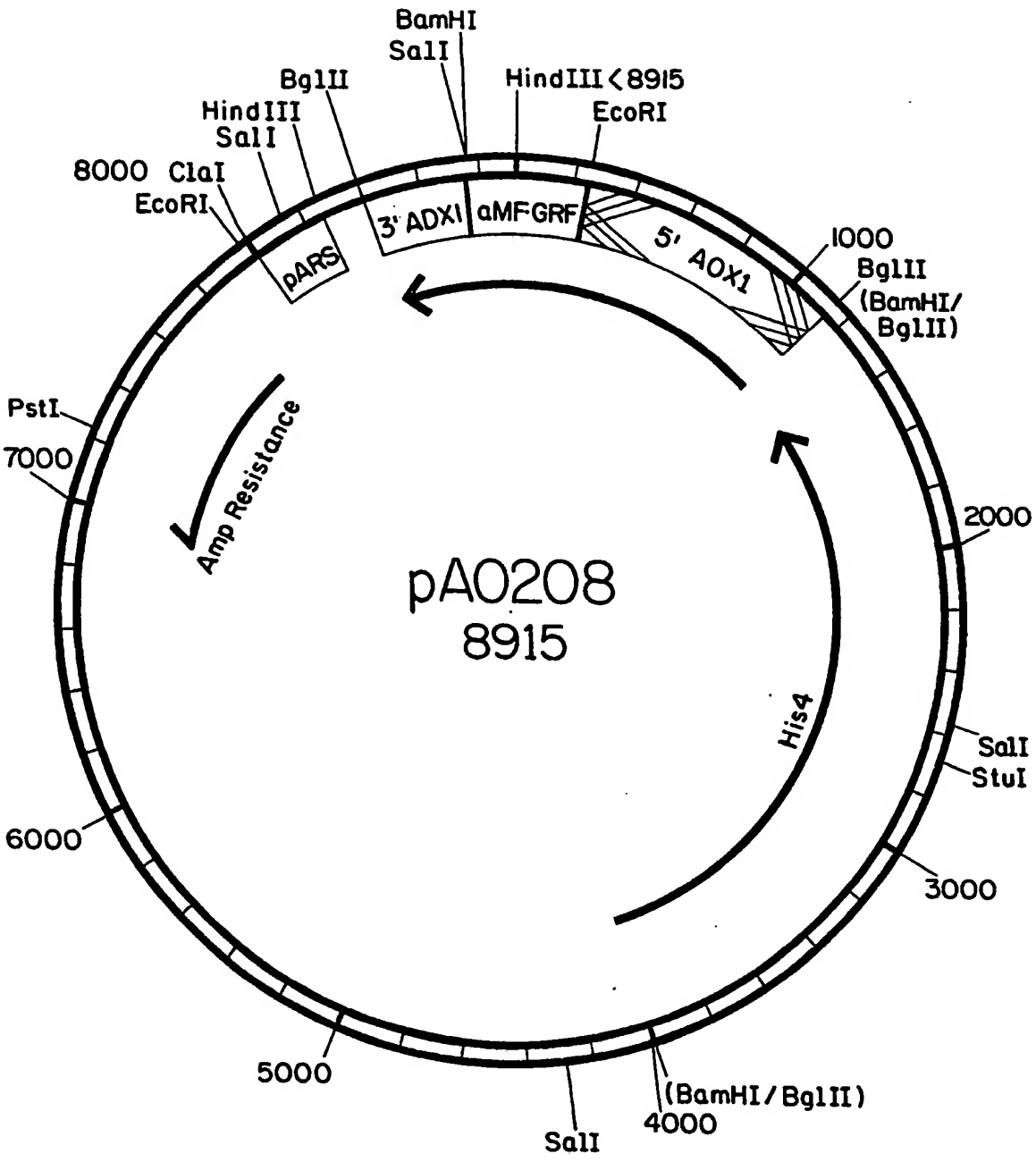
BamHI
SalI
BglII
HindIII
SalI
8000 ClaI
EcoRI
HindIII < 8915
EcoRI
3' ADX1
αMFGRF
5' AOX1
1000
BglII
(BamHI/
BglII)
pARS
PstI
7000
Amp Resistance
pA0208
8915
His4
2000
SalI
StuI
3000
6000
5000
SalI
(BamHI / BglII)
4000

FIG. 3

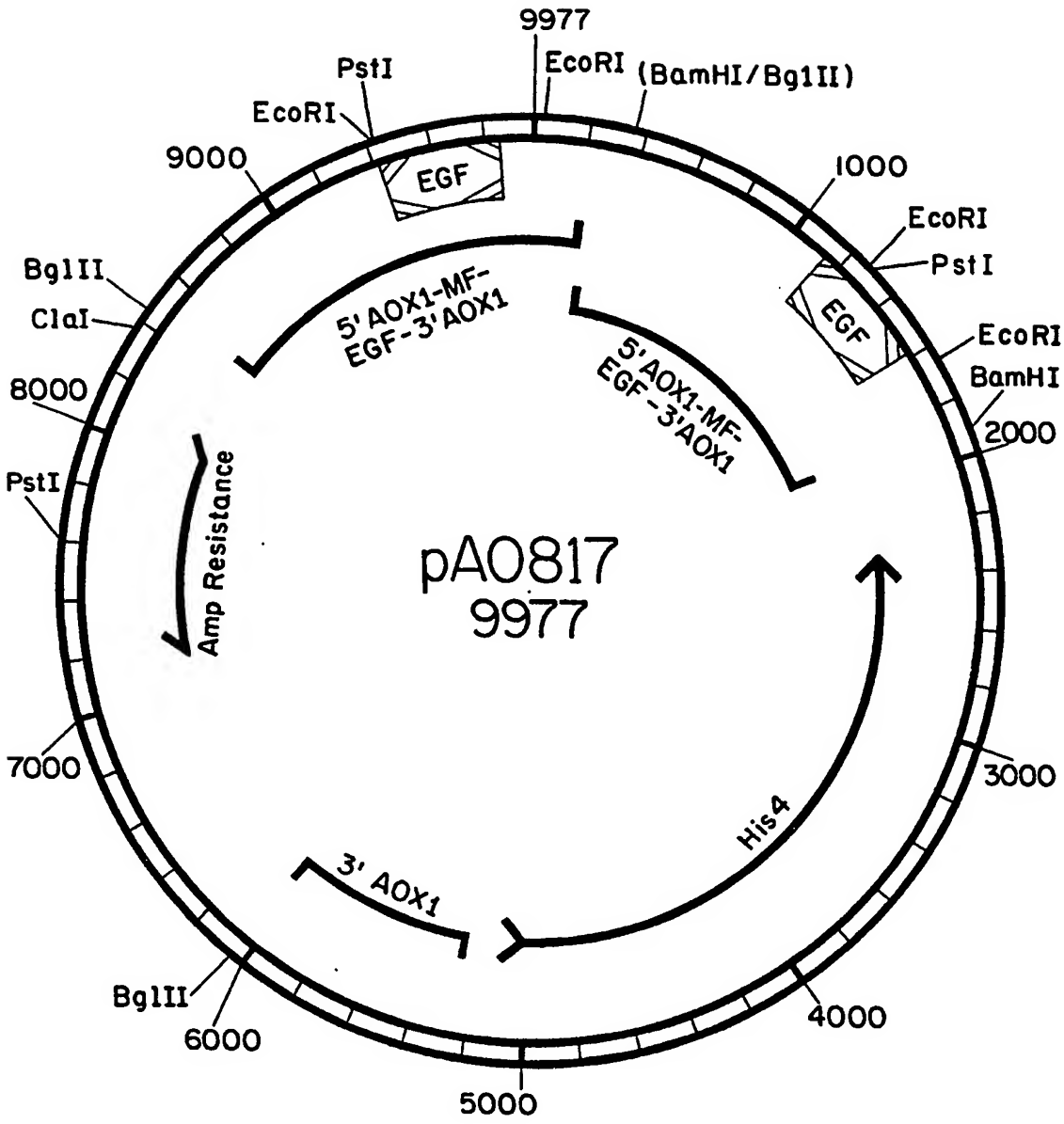
9977
EcoRI
(BamHI/Bg1II)
PstI
EcoRI
9000
EGF
1000
EcoRI
PstI
EGF
EcoRI
BamHI
2000
BglII
ClaI
8000
PstI
5' AOX1-MF-
EGF-3' AOX1
5' AOX1-MF-
EGF-3' AOX1
Amp Resistance
pA0817
9977
7000
3000
His4
3' AOX1
BglII
6000
4000
5000

FIG. 4

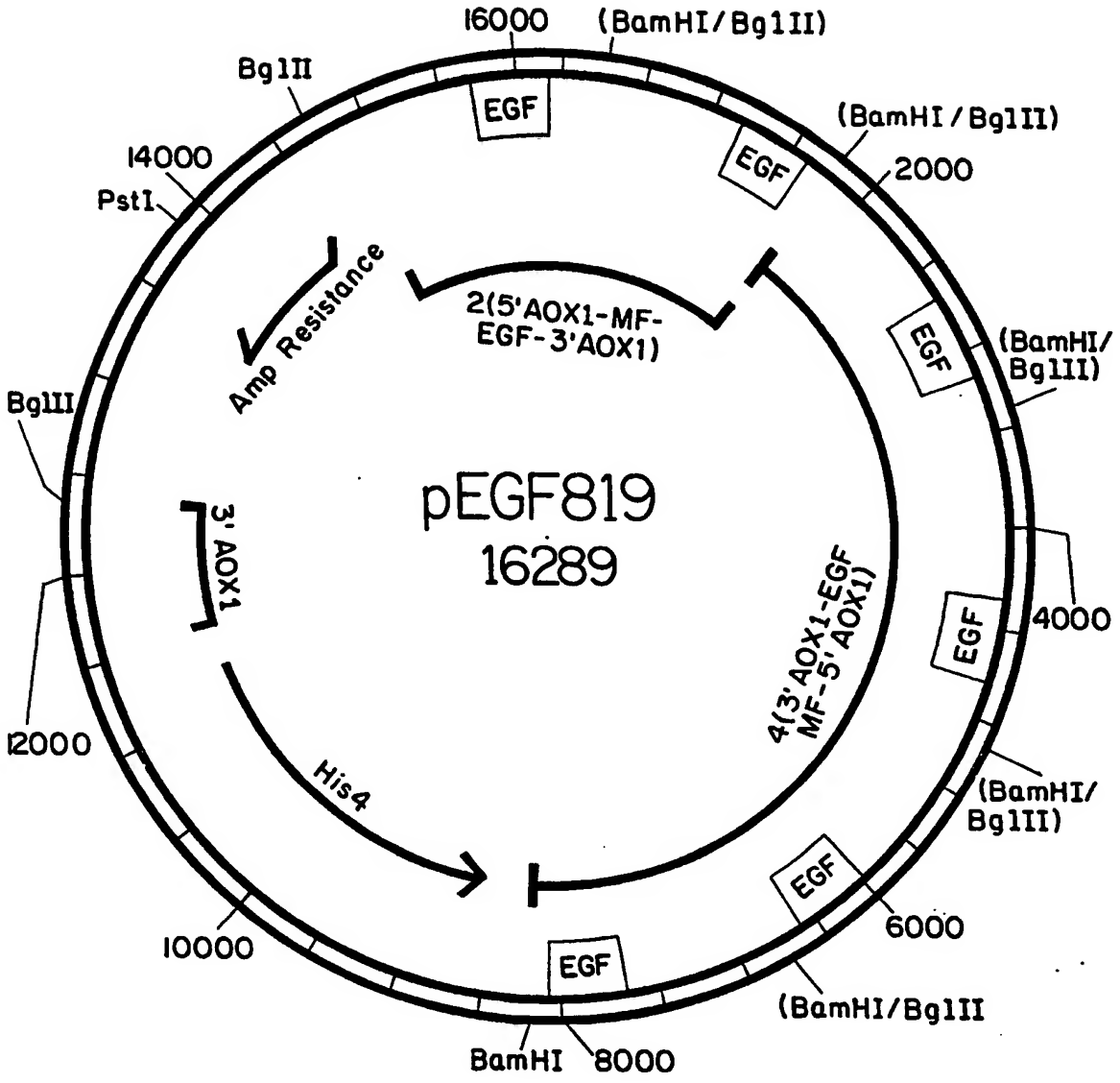
16000
(BamHI/Bg1II)
Bg1II
EGF
(BamHI/Bg1II)
14000
2000
PstI
EGF
Amp Resistance
2(5'AOX1-MF-
EGF-3'AOX1)
EGF
(BamHI/
Bg1II)
BglII
pEGF819
16289
3' AOX1
4(3'AOX1-EGF
MF-5'AOX1)
EGF
4000
12000
His4
(BamHI/
Bg1II)
EGF
6000
10000
EGF
(BamHI/Bg1II
BamHI
8000

FIG. 5

INTERNATIONAL SEARCH REPORT

International Application No PCT/US90/01353

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) 3

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC (5): C12N 15/00, C12P 21/00, C12P 21/02
US 435/69.3, 69.1, 255, 256, 320, 172.3

II. FIELDS SEARCHED

Minimum Documentation Searched 4

Classification System	Classification Symbols
U.S.	435/69.3, 69.1, 172.3, 255, 256, 320 536/27, 935/37, 28

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched 5

Computor Search Chemical Abstracts, Biological Abstracts: epidermal growth factor, Pichia, methanol inducible, yeast alpha mating factor, mut phenotype

III. DOCUMENTS CONSIDERED TO BE RELEVANT 14

Category *	Citation of Document, 16 with indication, where appropriate, of the relevant passages 17	Relevant to Claim No. 18
Y	Proceedings National Academy of Science Vol. 80 pp 7461-7465 issued December 1983, Urdea et al. "Chemical Synthesis of a Gene for Human Epidermal Growth Factor Urogastrone and its Expression in Yeast" See the entire document.	1-35
Y	Proceedings National Academy of Science Vol. 81, pp 4642-4646, issued August 1984, Brake et al. "∝- Factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae" See the entire document.	1-35
Y	Developments in Industrial Microbiology Vol. 29, pp 59-65, issued March 1988 Digan et al. "Secretion of heterologous proteins from the methylotrophic yeast, Pichia pastoris" See the entire document.	1-35
Y	Biotechnology Vol. 5 pp 1305-1308 issued December 1987, Tschopp et al. "high-level secretion of glycosylated invertase in the methylotrophic yeast, Pichia pastoris" See the entire document.	1-35

continued

* Special categories of cited documents: 15

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search 2	Date of Mailing of this International Search Report 2
30 MAY 1990	09 JUL 1990
International Searching Authority 1	**Signature of Authorized Officer 20**
ISA/U.S.	for Robin Teskin NGUYEN NGOC-HO INTERNATIONAL DIVISION

Form PCT/ISA/210 (second sheet) (May 1986)

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

Y	Biotechnology Vol. 5, pp 479-485 issued December 1987, Cregg et al. "High-level expression and efficient assembly of hepatitis B surface antigen in the methylotrophic yeast, Pichia pastoris" See the entire document.	1-35

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE [1]

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers .... because they relate to subject matter [1] not required to be searched by this Authority, namely:

2. ☐ Claim numbers....., because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out [1], specifically:

3. ☐ Claim numbers_____, because they are dependent claims not drafted in accordance with the second and third sentences of PCT Rule 6.4(a).

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING [2]

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.

2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

☐ The additional search fees were accompanied by applicant's protest.

☐ No protest accompanied the payment of additional search fees.